Niveau(x) d'enseignement: 7 Numéro EDRS : 3478123
Centre d'intérêt: 570

Brunet, Richard.

La précision de la mesure dans le questionnaire PERPE / Richard Brunet. -- [Montréal] : INRS-Education, 1973.
30, [10] f. : 4 graph. ; 28 cm.

Document: 1973-04. Bibliogr.: f. [40].

Descripteur(s): *Evaluation de l'enseignant *Evaluation par l'étudiant *Satisfaction de l'étudiant *Attitude de l'enseignant *Relations maître-élèves *Mesure *Test *Evaluation de tests *Questionnaire *Collégial (niveau)

Identificateur(s) de nom: *PERPE (Test)

L'étude de la fidélité d'un instrument consiste à définir le plus clairement possible ce qui doit être considéré comme erreur de mesure, le coefficient de fidélité n'étant par un but en soi mais plutôt une façon économique de résumer l'information. On peut distinguer deux grands types de coefficients de fidélité correspondant à deux définitions distinctes de l'erreur de mesure; ce sont le coefficient de précision et le coefficient de stabilité. Dans le cas du questionnaire PERPE (Perceptions étudiantes de la relation professeur-étudiants), l'étude du coefficient de stabilité serait prématurée à cette étape; c'est pourquoi cette recherche se limite à un examen de la précision de la mesure recueillie.

Cette étude de l'erreur de mesure dans le questionnaire PERPE se subdivise en cinq parties. Dans un premier temps, l'auteur identifie les sources principales de variation des scores; puis, il présente la méthodologie jugée appropriée pour jauger la part de la variance attribuable à l'erreur de mesure. La troisième partie présente et discute les résultats obtenus tandis que la quatrième partie traite de la stabilité des coefficients obtenus en les comparant à un deuxième groupe de coefficients issus d'un échantillon différent. Enfin, une cinquième partie présente brièvement quelques possibilités d'augmenter la fidélité des mesures.

Le schème d'analyse de variance simple "classes x étudiants" a été appliqué à 1196 classes dans lesquelles le questionnaire PERPE a été administré à l'automne 1970. La fidélité moyenne des item est de 0,90 pour les mesures de réalité, 0,68 pour les mesures d'attente, 0,87 pour les mesures d'insatisfaction brute, 0,82 pour les mesures d'insatisfaction nette et 0,76 pour le pourcentage des étudiants insatisfaits, une fois éliminés douze item dont la fidélité est jugée trop faible. La fidélité des scores composés est acceptable, telle que mesurée par le coefficient alpha de Cronbach et par la technique du "split-half" sur les classes. Enfin, les coefficients de fidélité obtenus sont stables, i.e. identiques sur deux échantillons différents de classes.

Type de document: r

EDUQ 8470129

I N R S - E d u c a t i o n

La précision de la mesure dans le questionnaire PERPE

Richard Brunet

30 octobre 1973

Document 1073-04

Le lecteur trouvera dans ce texte les informations pertinentes quant à la précision des mesures du questionnaire PERPE.

Etant donné que le schème d'analyse utilisé résulte d'un cheminement long et ardu, les étapes intermédiaires de ce cheminement feront l'objet des appendices A et B.

Table des matières

	pages
Préambule	
I - L'erreur de mesure dans le questionnaire PERPE	3
A. La fidélité des items du questionnaire PERPE	4
1. Sources de variation dans le test PERPE	4
2. Définition de l'erreur de mesure dans PERPE	5
3. Les mesures du questionnaire PERPE	7
4. Quelques techniques usuelles	8
B. Les scores composés	9
II - Méthodologie	10
A. Les techniques utilisées	11
1. Les items	11
2. Les scores composés	12
B. Les données	14
1. Les échantillons	14
2. Les contrôles effectués	14
III - Résultats	15
1. Les items	16
2. Les sections	22
IV - Stabilité des coefficients de précision	24
V - Nouveaux contrôles et implications quant à la fidélité des mesures	28
A. Contrôles actuels	29
B. Contrôles possibles	29
1. La sévérité différentielle	30
2. Les caractéristiques "sociologiques"	30
C. Conclusion	30
VI - Résumé et conclusion	31
Appendice A: fidélité intra-profil	
Appendice B: fidélité inter-profil	
Références	

Préambule

Avant de se lancer dans une étude sur la fidélité du test PERPE, il nous apparaît utile d'une part de préciser la notion de fidélité et d'autre part, de justifier l'exclusion de cette étude d'une partie des mesures fournies par l'instrument analysé, à savoir les normes.

. Notion de fidélité

La fidélité d'un instrument a plus souvent qu'autrement été définie par la technique utilisée pour calculer les coefficients présentés, ce qui a eu pour effet de faire perdre de vue le but réel d'une étude de la fidélité, à savoir la répartition de la variance totale en variance acceptable en fonction des buts de l'instrument (variance vraie) et en variance résiduelle considérée comme erreur de mesure.

Dans cet optique, l'étude de la fidélité d'un instrument consiste à définir le plus clairement possible ce qui doit être considéré comme erreur de mesure, le coefficient de fidélité n'étant pas un but en soi mais plutôt une façon économique de résumer l'information. De plus, ce coefficient aura une valeur en autant que la définition qu'on aura fournie de l'erreur de mesure sera adéquate.

On peut distinguer deux grands types de coefficients de fidélité correspondant à deux définitions distinctes de l'erreur de mesure; ce sont le coefficient de précision et le coefficient de stabilité (Lord et Novick, 1968, pp. 134-137). Dans le premier cas, l'indice obtenu reflète la reproductibilité des résultats hic et nunc; l'erreur de mesure n'inclut que l'infidélité de la procédure à laquelle la mesure est recueillie. Dans le second cas, l'erreur de mesure inclut tout changement chez l'examiné par rapport au trait mesuré sur une certaine période de temps. On voit qu'il y a autant de coefficients de précision pour un instrument donné que de conditions expérimentales permettant de recueillir les données.

Dans le cas du questionnaire PERPE, il nous apparaît prématuré, à ce stade des recherches, de parler d'un coefficient de précision. En effet, l'in-

clusion dans la variance d'erreur des changements temporels chez les professeurs implique d'une part des justifications théoriques de sorte qu'il soit possible de porter des conclusions sur l'instrument et d'autre part des contrôles expérimentaux passablement complexes. C'est pourquoi cette étude se limitera à un examen de la précision de la mesure recueillie par le questionnaire PERPE.

2. Les normes

Les résultats fournis par le questionnaire PERPE aux professeurs sont de deux ordres, à savoir des scores obtenus et des scores permettant au professeur de se comparer à ses collègues (normes).

Etant donné d'une part cette différence de niveau entre les deux types de mesure et d'autre part la non-redondance[1] des indices de fidélité d'un type de score sur l'autre, il nous apparaît préférable de traiter ce sujet dans une étude différente de celle-ci.

Cette étude de l'erreur de mesure dans le questionnaire PERPE se subdivise en cinq parties. Dans un premier temps, nous identifierons les sources principales de variation des scores; puis, nous présenterons la méthodologie que nous jugeons appropriée pour jauger la part de la variance attribuable à l'erreur de mesure. La troisième partie présentera et discutera les résultats obtenus tandis que la quatrième partie traitera de la stabilité des coefficients obtenus en les comparant à un deuxième groupe de coefficients issus d'un échantillon différent. Enfin, dans un cinquième temps, nous présenterons brièvement quelques possibilités d'augmenter la fidélité de nos mesures.

---

[1] Comme Lord et Novick (1968, p. 215) le mentionnent, les normes sont une transformation non-linéaire des mesures originales; par conséquent, la fidélité de ces mesures ne peut être transposée aux mesures normalisées.

- I -

L'erreur de mesure
dans le questionnaire PEPPE

Le questionnaire PERPE est un instrument servant à évaluer différents aspects de la situation éducative où sont mis en présence un groupe d'étudiants et un professeur. Suite à l'administration de PERPE, ce professeur reçoit un dossier constitué des moyennes des réponses A et B (MA et MB) fournies par les étudiants. De plus, trois scores sont calculés pour chaque item, à savoir les MAB, les MAL et les %; ces scores réfèrent aux moyennes absolues et algébriques des différences entre les réponses A et les réponses B et au pourcentage d'étudiants insatisfaits dans la classe. Enfin, le portrait présente des scores de sections et des scores globaux relatifs à l'insatisfaction brute et nette et aux %; ces scores sont formés à partir d'une combinaison linéaire des scores aux items qui composent la section ou le test.

Le terme "résultats au questionnaire PERPE" fait alors référence au dossier du professeur et non pas aux réponses des étudiants prises individuellement. Il importe également d'introduire une distinction entre les deux types de résultats fournis par le test PERPE, à savoir les résultats à chacun des items et les scores composés (de section ou global), ces derniers étant une combinaison linéaire des résultats aux items. En effet, ces deux types de scores relèvent de deux bases théoriques différentes et c'est pourquoi nous les discuterons séparément.

A. La fidélité des items du questionnaire PERPE

Avant de définir une méthodologie qui permettra d'évaluer la fidélité des items du questionnaire PERPE, il importe de préciser les diverses sources de variation des scores à ce questionnaire.

1. Sources de variation dans le test PERPE

On peut identifier trois sources principales de variation dans le questionnaire PERPE. Ces trois composantes sont attribuées aux professeurs (ou à la situation mesurée), aux étudiants et à l'erreur de mesure non prédictible (non systématique).

a) La composante "situation" (professeur)

Chaque item de PERPE veut mesurer un comportement ou une attitude du professeur en classe. Il est donc légitime de prétendre que ce comportement puisse être un des déterminants du résultat que recevra ce professeur. Cette composante peut être vue sous deux aspects, à savoir l'aspect immédiat et l'aspect temporel. Dans le premier cas, l'intérêt porte sur la situation mesurée à un moment précis dans le temps et dans l'espace alors que, dans le second cas, l'objet d'étude est plutôt l'évolution ou le changement des comportements.

b) La composante "étudiant"

Le comportement ou l'attitude "réelle" du professeur nous est connu via la perception qu'en ont les étudiants; par conséquent, le jugement porté par chacun des étudiants est fonction du schème de référence de chacun des étudiants. Cette composante a également deux aspects (immédiat et temporel) du même type que pour la composante "situation mesurée". Quelques exemples de la composante "étudiant" sont la sévérité différentielle des étudiants face à l'objet mesuré, et l'influence sur les réponses de variables indépendantes telles le sexe, l'âge, le niveau socio-économique, etc...

c) La composante "résiduelle"

Une dernière source de variation consiste en l'erreur de mesure non prédictible, non systématique et en toute autre composante non définie précédemment.

2. Définition de l'erreur de mesure dans PERPE

L'utilisation d'une technique particulière d'estimation de la fidélité suppose auparavant que des décisions aient été prises quant à ce qui doit être considéré comme erreur de mesure. Ces décisions conditionneront l'interprétation qui sera donnée des coefficients obtenus. De plus, la définition de l'erreur de mesure doit se faire en conformité avec l'utilisation pratique qui est faite de l'instrument. Ainsi, le professeur qui utilise PERPE n'a aucun contrôle sur la composition du groupe de juges; par conséquent, il nous apparaît essentiel d'inclure dans l'erreur de mesure la composante "étudiant immédiat". En d'autres mots, toute variation entre les étudiants face à l'évaluation du professeur sera traitée comme variance d'erreur. Cette

définition permettra de voir jusqu'à quel point un professeur obtiendrait des résultats identiques s'il faisait passer une seconde fois le test à un groupe comparable d'étudiants sans avoir modifié son comportement.

La composante "étudiant temporel" ne peut actuellement être évaluée puisqu'aucune donnée n'est disponible. Il est par conséquent, impossible de savoir si un professeur ne modifiant pas son comportement obtiendrait, avec les mêmes étudiants, des résultats équivalents lors d'une seconde passation du test, les différences observées étant alors attribuables à des caractéristiques telles l'humeur au moment de la passation, etc. Toutefois, il semble reconnu que ce type de variation ne revêt pas une grande importance.

La composante résiduelle est évidemment incluse dans l'erreur puisque non systématique.

L'aspect temporel de la composante "situation-professeur" ne sera pas touché dans cette étude; en effet, tel que mentionné précédemment, cette composante n'est que très difficilement évaluable (par un test-retest) et les résultats fournis par une étude de ce type seraient difficilement interprétables puisque les composantes intervenant entre le test et le retest sont nombreuses et ne doivent pas toutes être assimilées à l'erreur de mesure. Ainsi, un coefficient de fidélité qui ne tiendra pas compte de cet aspect nous renseignera sur l'évaluation par les étudiants de comportements situationnels, hic et nunc. En d'autres mots, des résultats fidèles nous renseigneront sur la reproductibilité des mesures de comportements prises à un moment donné et dans un espace donné.

L'aspect immédiat de la composante "situation-professeur" nous apparaît être le seul à devoir être inclus dans la variance vraie. Par conséquent, tout déterminant des résultats qui n'est pas le comportement du professeur ou la situation mesurée tels que perçus au moment où le test est administré à un groupe d'étudiants donné, est considéré comme source d'erreur de mesure. Pour employer la terminologie de Lord et Novick (1968)[1], la fidélité du test PERPE réfère à son coefficient de précision et non à un coefficient de stabilité.

---

[1] voir pp. 134 et sq.

3. Les mesures du questionnaire PERPE

Les diverses sources de variation identifiées précédemment jouent de façon différente, selon qu'on parle des mesures de réalité, d'attente ou d'insatisfaction. Nous essaierons, dans les quelques lignes qui suivent, de spécifier un peu plus clairement les déterminants de chaque mesure.

a) Scores MA

Pour un professeur donné, la MA à un item est la moyenne des perceptions des étudiants quant à la réalité mesurée par cet item. Cette moyenne est donc déterminée d'une part par la réalité "objective" qu'on veut mesurer et par la perception différentielle des juges de cette réalité.

b) Scores MB

Pour un professeur donné, la MB d'un item est la moyenne des attentes des étudiants face au professeur ou à la situation évalués. Ce score est donc déterminé d'une part par la réalité mesurée et d'autre part par la composante étudiante. Cette dernière composante devrait logiquement avoir un effet beaucoup plus important dans le cas des MB que pour les MA puisque les attentes sont beaucoup plus personnelles qu'un comportement objectif.

Les MB sont également déterminées jusqu'à un certain point par la désirabilité sociale du trait mesuré. En effet pour la majorité des items du test PERPE, la présence du trait mesuré est hautement désirable alors que son absence est hautement indésirable. Ceci a pour effet d'obliger les MB à être supérieures aux MA, réduisant ainsi leur variance.

Une dernière composante, plutôt technique, détermine jusqu'à un certain point les MB. Il s'agit en occurence de la forme des échelles utilisées; en effet, les MA se situent habituellement autour du point 3 de l'échelle à 5 points, ce qui ne laisse que peu de place à la MB pour varier puisque dans la plupart des cas, cette MB doit être supérieure à la MA. Si l'échelle de réponses utilisée avait été plus raffinée, il est possible que les MB aient pu avoir une variance appréciablement plus grande.

En résumé, il est possible de formuler l'hypothèse que la fidélité des MB sera plus faible que celle des MA à cause d'une part de la variabilité

plus petite de ces scores et d'autre part du fait que la composante "étudiant", plus importante dans le cas des MB que dans celui des MA, est placée dans la variance d'erreur.

c) Scores d'insatisfaction

Les scores d'insatisfaction calculés dans le profil au test PERPE sont des scores dérivés des mesures de réalité et d'attente et sont par conséquent, soumis aux mêmes déterminants que ceux-ci. En effet, ces scores d'insatisfaction seront fidèles en autant qu'ils représentent la différence entre la composante situation telle que perçue dans la réalité et telle qu'attendue. Donc, toute variance qui ne peut être attribuée au facteur "situation mesurée" devra être incluse dans la variance d'erreur.

En termes statistiques la fidélité des scores d'insatisfaction sera proportionnelle à la fidélité des MA et des MB et inversement proportionnelle à la corrélation entre ces deux mesures (Stanley, 1971). Connaissant la forte relation entre les MA et les MB (médiane de 0.68), on peut s'attendre à ce que, en règle générale, la fidélité des mesures d'insatisfaction soit inférieure à celle des MA.

4. Quelques techniques usuelles

On rencontre, dans le domaine des tests d'aptitudes, quelques techniques standards pour l'évaluation de la fidélité d'un instrument. Chaque technique a sa définition particulière de l'erreur de mesure. Nous essaierons dans cette partie de justifier l'emploi de la technique que nous privilégions par rapport aux autres techniques disponibles.

a) Test-retest

Cette technique n'est pas adaptée à la situation du test PERPE puisque un retest implique nécessairement la présence indissociable des composantes temporelles, l'une due aux répondants, l'autre à l'évolution de la situation mesurée, cette dernière ne devant pas selon nous être considérée comme erreur de mesure. En effet, s'il est raisonnable dans le domaine des aptitudes de postuler que le trait mesuré est invariant chez chaque personne mesurée sur une période de temps assez longue pour que la mémoire ne joue pas, les différences étant alors attribuables aux répondants, il nous appa-

raît peu pertinent de postuler, dans le cas de PERPE, qu'il ne puisse y avoir évolution dans le temps de l'objet mesuré. De plus, cette technique ne tient pas compte de la source d'erreur due à l'échantillonnage étudiant. Par conséquent, cette technique fournirait un estimé très inexact de la précision réelle de chaque item.

b) "Split-half" sur les juges ou analyse de variance des juges

Ces deux techniques ont été regroupées sous la même rubrique à cause de leur lien étroit de parenté; en effet, les deux méthodologies adoptent la même définition de l'erreur de mesure, définition qui est identique à celle que nous avons formulée précédemment (voir section 2). De plus, l'analyse de variance fournit un indice de fidélité qui est la moyenne des coefficients obtenus en faisant tous les "split half" possibles. C'est pourquoi nous privilégions cette méthodologie pour l'estimation de la fidélité des items.

b. Les scores composés

Le test PERPE fournit aux professeurs des scores d'insatisfaction à six sections et des scores globaux d'insatisfaction, chacun de ces scores étant la moyenne des scores à plusieurs items.

On retrouve, sous-jacents à la construction des scores de section (ou du score global), deux construits, le premier étant l'insatisfaction et le second étant un trait commun à une série d'items définis par l'appellation donnée aux sections (ou par le score global). Il serait hors de propos de discuter ici de la justesse des regroupements qui ont été effectués; qu'il suffise de mentionner que ces regroupements sont basés sur une certaine homogénéité des items (Gagné, 1970; Hamel, 1972).

Par conséquent, le problème de l'évaluation de la fidélité de ces scores est identique à celui qu'on retrouve dans le domaine des tests d'aptitudes, problème qui trouve sa solution par la technique du alpha de Cronbach. Le rationel de cette technique classique est très bien expliqué par Nunnally (1967).

Méthodologie

Nous préciserons, dans un premier temps, les techniques utilisées pour le calcul des coefficients de fidélité et, dans un second temps, les données qui ont servi de base aux calculs.

A. Les techniques utilisées

Trois techniques correspondant à deux bases théoriques différentes seront utilisées, la première applicable aux items et les deux autres aux scores composés.

1. Les items

Soit, pour chacun des 61 items et pour chaque mesure, le tableau suivant:

classes (i)

étudiants (j)	1	2	3	4		i		K
1	$X_{1,1}$			$X_{4,1}$				$X_{K,j}$
2	$X_{1,2}$			$X_{4,2}$		$X_{i,j}$		
3	$X_{1,3}$			$X_{4,3}$				
N_i	X_{1,n_i}			X_{4,n_i}		X_{i,n_i}		X_{K,n_i}

L'analyse de variance permet de partitionner la variance totale en variance due aux classes et en variance résiduelle, ce qui correspond exactement à la définition que nous avons donnée de la variance vraie (variance due aux classes) et de l'erreur de mesure (variance résiduelle).

Le coefficient de fidélité d'un étudiant est obtenu par la formule suivante:

$$F = \frac{CM_c - CM_e}{CM_c + (L - 1)\,CM_e}$$

k = Nombre

correspond à ? ou à $R^2_{x,x'}$??

$$\text{où} \quad CM_c' = \frac{\sum\sum X_{ij}^2 - \sum \frac{(\sum X_{ij})^2}{n_i}}{\sum n_i - K}$$

$$\text{où} \quad CM_e = \frac{\sum \frac{(\sum X_{ij})^2}{n_i} - \frac{(\sum\sum X_{ij})^2}{\sum n_i}}{K - 1}$$

$$\text{et où} \quad L = (\sum n_i - \sum n_i^2 / \sum n_i) / (K - 1)$$

à savoir une correction afin de tenir compte de l'inégalité du nombre d'observations dans chaque classe. La fidélité d'un répondant peut ensuite être extrapolée à une classe de n étudiants à l'aide de la formule de Spearman-Brown:

$$r^2_{n,n} = \frac{n\, r^2_{1,1}}{1 + (n - 1)\, r^2_{1,1}}$$

où $r^2_{1,1}$ est la fidélité d'un répondant,

$r^2_{n,n}$ est la fidélité de n répondants

et n est égal à 27, à savoir le nombre moyen d'étudiants par classe.

Il faut noter que, dans le cas des pourcentages d'étudiants insatisfaits, les nombres bruts plutôt que les pourcentages seront utilisés.

Le lecteur intéressé aura sans doute remarqué qu'il s'agit ici d'un schème d'analyse de variance simple (nombres inégaux) dont Winer (1971, pp. 283-300) et Stanley (1971, pp. 420 sq) expliquent le rationnel.

2. Les scores composés

Les scores composés étant fondés sur l'homogénéité du contenu mesuré et/ou du type de mesure effectué, le problème de la détermination des coefficients de fidélité de ces scores peut être abordé de deux façons,

à savoir en fonction de la consistance interne des items entrant dans la composition de ces scores (coefficient alpha de Cronbach) et en fonction de la définition de l'erreur de mesure fournie précédemment. Ces deux aspects donnent lieu à deux techniques d'analyse, à savoir le coefficient alpha de Cronbach et la corrélation inter-profil.

a) Alpha de Cronbach

Ce coefficient dont la formule est la suivante

$$\alpha = \frac{K}{K-1} \cdot \left(1 - \frac{\Sigma \sigma_i^2}{\sigma_t^2}\right)$$

évalue la représentativité des items utilisés pour mesurer le construit sous-jacent. Cette représentativité est définie d'une part par le degré de relation entre les divers items formant le score composé et, d'autre part par le nombre d'items utilisés pour effectuer cette mesure. Le lecteur trouvera une description détaillée du rationnel de cette technique dans Nunnally (1967, pp. 194-198). Ce coefficient fournit alors un estimé de la reproductibilité de ces scores de section si on utilisait d'autres items issus de la même banque d'items. Par conséquent, ce coefficient sera alors une limite inférieure de la reproductibilité réelle de ces scores puisque les mêmes items sont utilisés pour les mêmes étudiants.

b) Corrélation inter-profil

Etant donné Z, l'ensemble des étudiants d'une classe, deux sous-groupes sont formés au hasard (X et Y). Deux séries de scores de section sont calculées, la première à partir des réponses du sous-groupe X et la seconde à partir de celles du sous-groupe Y. Cette procédure est appliquée à chaque classe de la population, de sorte que, pour chaque classe, deux MAL et deux MAB de section calculées sur des sous-groupes différents sont disponibles.

Pour chaque section, la corrélation corrigée par la formule de Spearman-Brown, entre les profils des demi-classes, constitue un indice de la fidélité de cette section. En effet, ce coefficient reflète jusqu'à quel point l'ordre des professeurs à une section demeure inchangé.

Dans un tel schème, la définition de la variance vraie est la même que celle que nous avons adoptée.

B. Les données

1. Les échantillons

Pour des fins d'étude de la stabilité des coefficients de fidélité, deux échantillons seront utilisés, le premier comprenant les 1196 classes auxquelles le test PERPE a été administré à l'automne 1970, et le second les 728 classes de l'administration du printemps 1971. Dans la suite du texte, nous nommerons ces échantillons A et B respectivement, les classes sont toutes du niveau collégial. La description détaillée des deux échantillons est fournie par Renaud (1973).

2. Les contrôles effectués

Le traitement mécanographique des réponses des étudiants prévoit certains contrôles, entre autre l'exclusion des calculs des étudiants marginalement insatisfaits et de ceux qui n'ont pas répondu à un nombre suffisant d'items (bernier et al., 1970, p. 5).

Comme ces étudiants sont exclus du dossier du professeur, ceux-ci seront aussi exclus des données de base servant au calcul des coefficients de fidélité.

Résultats

Nous examinerons dans un premier temps les résultats relatifs à chaque item pour passer ensuite aux résultats relatifs aux sections. Le tableau 1 contient pour chaque item le coefficient de fidélité de chaque mesure tandis que le tableau 3 présente les indices de fidélité des scores composés quant à l'insatisfaction.

1. Les items

Les coefficients de fidélité de chaque item pour chaque mesure font l'objet du tableau 1. A cause des différences marquées entre la fidélité de chaque mesure, différences reliées à la définition de l'erreur de mesure, nous examinerons une à une les différentes mesures du test.

a) La mesure de la réalité (MA)

La MA d'un professeur est la moyenne des perceptions des étudiants face à la réalité mesurée. Comme la méthodologie utilisée définit la variance d'erreur par tout ce qui n'est pas dû au professeur, le coefficient de fidélité obtenu permet de vérifier plusieurs points tels la compréhension univoque de l'item par les étudiants et la cohésion intra-classe par rapport à la dispersion inter-classe.

L'examen de tout coefficient de fidélité pose cependant le problème du seuil acceptable pour un tel coefficient. Ce seuil dépend d'une part de l'utilisation qui sera faite des résultats au test et d'autre part de la nature des choses mesurées. Dans le cas du questionnaire PERPE, les aspects qui y sont mesurés ne peuvent l'être aussi précisément que dans le cas de tests d'aptitudes où le coefficient de fidélité atteint très souvent 0.9%. De plus, le questionnaire PERPE est confidentiel et ne sert pas à la sélection de professeurs; il n'est pas utilisé dans un but normatif. C'est pourquoi un seuil de 0.80 nous apparaît acceptable dans le cas présent. Le lecteur trouvera d'autres justifications à l'emploi d'un tel seuil en examinant les divers seuils utilisés par d'autres instruments de même type, quoique la comparaison aux autres instruments ne soit pas en soi un critère.

Tableau 1

Fidélité des items par ANOVA, échantillon A

	MA	MB	MAL	MAR	%
1	.94	.84	.88	.80	.76
2	.82	.82	.81	.76	.70
3	.85	.59	.78	.84	.76
4	.85	.53	.81	.82	.73
5	.90	.75	.86	.85	.78
6	.86	.59	.81	.82	.76
7	.80	.56	.75	.77	.68
8	.82	.63	.75	.75	.69
9	.94	.80	.85	.76	.69
10	.91	.80	.71	.68	.62
11	.85	.76	.84	.62	.58
12	.90	.81	.77	.73	.68
13	.90	.80	.77	.74	.68
14	.89	.72	.86	.80	.71
15	.74	.58	.59	.74	.69
16	.91	.69	.89	.84	.78
17	.88	.73	.78	.77	.71
18	.91	.78	.85	.82	.76
19	.95	.88	.86	.66	.60
20	.97	.90	.91	.84	.79
21	.98	.94	.93	.93	.90
22	.90	.70	.82	.79	.71
23	.88	.62	.83	.79	.71
24	.92	.84	.82	.80	.74
25	.90	.76	.83	.85	.79
26	.92	.79	.85	.77	.72
27	.89	.66	.80	.80	.74
28	.91	.85	.90	.86	.84
29	.95	.59	.93	.93	.89
30	.91	.58	.88	.88	.82
31	.94	.89	.92	.92	.87
32	.95	.73	.90	.89	.86
33	.93	.73	.87	.87	.81
34	.88	.61	.83	.84	.75
35	.87	.89	.82	.85	.78
36	.86	.54	.84	.86	.77
37	.81	.60	.78	.77	.67
38	.90	.64	.88	.84	.75
39	.87	.51	.86	.86	.77
40	.85	.50	.82	.82	.70
41	.84	.56	.79	.79	.71
42	.86	.59	.79	.80	.73
43	.87	.88	.83	.85	.78
44	.81	.55	.71	.78	.64
45	.88	.70	.76	.78	.66
46	.82	.55	.75	.76	.66
47	.89	.64	.84	.82	.76
48	.78	.63	.70	.75	.64
49	.84	.57	.77	.80	.73
50	.87	.49	.84	.84	.79
51	.81	.51	.76	.77	.67
52	.78	.58	.80	.81	.81
53	.91	.72	.87	.81	.82
54	.88	.72	.75	.78	.71
55	.87	.73	.81	.77	.69
56	.93	.82	.81	.61	.55
57	.84	.61	.85	.82	.77
58	.87	.70	.77	.75	.71
59	.81	.66	.65	.65	.62
60	.87	.72	.78	.76	.71
61	.88	.58	.75	.74	.69

L'examen des résultats relatifs aux MA (voir tableau 1) nous amène à identifier trois items dont la fidélité est inférieure à 0.80; il s'agit des items 15, 48 et 52. De plus, 14 items ont un coefficient inclus dans l'intervalle 0.80-0.85; ce sont les items 02, 03, 04, 07, 08, 11, 37, 40, 41, 44, 46, 49, 51 et 61.

En somme, face au seuil de 0.80 fixé préalablement, seulement trois items ne répondent pas aux normes fixées; de ce nombre, un item (question 15) devrait être éliminé à cause d'ambiguités logiques qui se reflètent d'ailleurs sur la fidélité puisque cet item est le moins fidèle des 61 items du test au niveau de la mesure de réalité.

b) La mesure d'attentes (MB)

La fidélité de la mesure d'attentes pose plusieurs problèmes, tel que vu précédemment. La définition de l'erreur de mesure posée au chapitre précédent est très restrictive en ce qu'elle considère tout ce qui n'est pas dû à la situation ou au professeur comme erreur de mesure; on peut cependant discuter de la logique sous-jacente à ce choix en ce sens que les attentes des étudiants dépendent à la fois de la situation extérieure à eux et de variables personnelles; il aurait alors fallu inclure dans la variance vraie la part de variance due aux étudiants. Cependant, dans la mesure où la MB est définie comme étant caractéristique de la situation (ou du professeur), ou, en d'autres mots, dans la mesure où les attentes moyennes des étudiants sont fournies au professeur, la variance due aux étudiants doit être classée dans l'erreur de mesure, et notre définition de l'erreur de mesure doit être considérée comme toute adéquate par rapport à la logique sous-jacente au test PERPE.

Comme la MB est une mesure recueillie directement auprès des étudiants et non une mesure dérivée, il nous apparaît juste de conserver pour cette mesure le seuil de 0.80 utilisé dans le cas des MA.

L'examen du tableau 1 révèle que seulement 10 items ont une fidélité supérieure au seuil fixé, alors que 13 items ont une fidélité supérieure à 0.75.

Ces resultats remettent en cause l'opportunité de fournir aux professeurs les attentes moyennes des étudiants. Trois hypothèses majeures peuvent expliquer l'impossibilité d'obtenir, dans le cas du test PERPE, des mesures d'attentes fidèles; la première consiste en l'ambiguité logique de la définition d'attentes utilisée par PERPE, où il a été remarqué que les étudiants répondaient parfois selon le professeur idéal et parfois selon des attentes "circonstanciées", ce qui contribuerait à augmenter indûment la variance intra-classe. La seconde hypothèse consiste en [illegible] de ce qui est mesuré; en effet il n'est peut-être pas possible de mesurer fidèlement les attentes des étudiants. Une troisième hypothèse, plus terre à terre mais qui a un certain lien avec l'hypothèse précédente, a trait à l'échelle de mesure: toutes les distributions des scores MB sont fortement asymétriques et leptokurtiques de sorte qu'une bonne discrimination est théoriquement à peu près impossible.

Avant de statuer définitivement sur la mesure d'attentes sur la seule base de la fidélité, il nous semble plus opportun d'étudier plus à fond la validité comparée de différentes méthodes de mesure, ainsi que leur fidélité respective. Entre-temps, les attentes telles que mesurées par le test PERPE ont une fidélité beaucoup trop faible à notre avis, ce qui remet en cause leur utilité pratique, sans toutefois nier l'intérêt théorique qu'une telle mesure puisse susciter.

c) Les mesures d'insatisfaction (MAL, MAB et %)

Il ne nous apparaît pas essentiel de traiter séparément de la fidélité des trois mesures d'insatisfaction puisque, comme on peut le constater au tableau 2, les corrélations entre ces mesures au niveau de la fidélité sont très élevées.

L'examen des fidélités de ces trois mesures d'insatisfaction présentées au tableau 1 révèle que la fidélité moyenne, après transformation en scores Z de Fisher, des MAL est sensiblement égale à celle des MAB alors que la fidélité des % se situe à un niveau inférieur. Cette baisse de la fidélité s'explique par la finesse de l'échelle de mesure utilisée; en effet, le pourcentage d'étudiants insatisfaits dans une classe est issu d'une échel-

Tableau 2

Corrélations entre les fidélités (transformées en Z de Fisher) des trois mesures d'insatisfaction (N = 61 items)

	MAB	%
MAL	79	81
MAB		97

le dichotomique de type insatisfait-non-insatisfait où sont considérés non-insatisfaits tous les étudiants dont la réponse A et la réponse B est identique alors que sont considérés insatisfaits tous les étudiants dont les réponses A et B diffèrent.

Il n'est donc pas surprenant d'observer une fidélité moyenne des % inférieure à celle des MAB puisque les premières ne constituent en partie qu'une réduction de l'information déjà contenue dans les secondes.

Avant d'aborder l'examen des items, il nous apparaît essentiel de clarifier trois points, à savoir le statut des scores d'insatisfaction, le critère de fidélité et l'asymétrie des distributions.

1) Les trois scores d'insatisfaction sont issus des réponses A et B fournies par les étudiants. Par voie de conséquence, ces scores cumuleront à la fois les avantages et les défauts des scores d'origine; par défaut, nous entendons l'infidélité relative des scores d'origine et leur inconsistence logique.

2) Compte tenu de ce fait, la fidélité des scores d'insatisfaction sera généralement inférieure à celle des MA. Il nous apparaît donc justifié dans ce cas d'abaisser le critère de fidélité à 0.75.

3) Il existe une tendance à peu près universelle dans toutes les mesures d'insatisfaction à observer des distributions fortement asymétriques des scores. Le test PERPE n'échappe pas à cette tendance; en effet, pour la majorité des items du questionnaire, les scores observés ont tendance à se regrouper autour du point 0.0

qui représente l'absence d'insatisfaction, de sorte que beaucoup de scores se situent sur une partie très restreinte du continuum. Ainsi, ces scores pourront être fort peu fidèles alors que les scores d'insatisfaction plus extrêmes pourront l'être. Concrètement, les scores inférieurs à la moyenne pourraient fort bien ne pas être fidèles alors que les scores supérieurs à cette moyenne pourraient être très fidèles. Les implications pratiques d'un tel état de fait sont importantes, notamment face à la normalisation. Notons toutefois que nous ne traiterons que de la fidélité d'un item en général, reportant à une étude ultérieure l'aspect de la fidélité différente d'un item à divers niveaux de l'échelle de mesure.

En se référant d'une part au critère de fidélité défini précédemment et d'autre part, aux résultats faisant l'objet du tableau 1, on remarque que les items 10, 11, 15, 44, 48, 59 et 61 ont une fidélité inférieure à 0.75 à chacune des trois mesures d'insatisfaction. Pour tous ces items, sauf pour l'item 10, la fidélité de la mesure de réalité est inférieure à la fidélité moyenne des MA.

On remarque de plus que les items 02, 12, 13, 19, 37, 45 et 56 ont une fidélité acceptable sur une seule des trois mesures d'insatisfaction. De ce nombre, les items 19 et 56 sont très particuliers en ce sens que la différence entre leur fidélité, au niveau de la MAL et celle au niveau de la MAB est nettement disproportionnée par rapport à la différence observée sur les autres items; nous n'avons pu trouver d'explication à ce phénomène. Compte tenu du fait que, pour ces deux items, la fidélité de la MA est excellente, alors que, pour les autres, elle se situe sous la moyenne, nous suggérons que ces items soient conservés.

Ainsi, 12 items devraient à notre avis être soit modifiés ou éliminés; ce sont les items 02, 10, 11, 12, 13, 15, 37, 44, 45, 48, 59 et 61. Si on exclut ces items du calcul de la fidélité moyenne pour chaque mesure, ces moyennes sont de 0.90, 0.68, 0.83, 0.82 et 0.76 respectivement pour les MA, MB, MAL, MAB et le pourcentage d'étudiants insatisfaits.

d) Résumé

L'étude de la fidélité des items du questionnaire PERFE a donné lieu à quelques observations qu'il nous apparaît utile de résumer:

1) la mesure de la réalité (MA) est en général très fidèle;
2) la mesure d'attentes semble très peu fidèle, du moins en fonction de la définition que nous avons donnée de la variance d'erreur;
3) les mesures d'insatisfaction brute et nette (MAB et MAL) ont un niveau de fidélité fort acceptable, sauf dans le cas de quelques items,
4) le pourcentage d'étudiants insatisfaits est en général peu fidèle. Il y aurait lieu à notre avis de remettre en cause l'utilité de cette mesure à la lumière de ces résultats d'une part et d'autre part de la relation qui unit cette mesure à la MAB;
5) nous suggérons l'élimination ou la modification de 12 items ne rencontrant pas le niveau de fidélité attendu.

2. Les sections

Tel que mentionné précédemment, deux techniques sont utilisées pour évaluer la fidélité des scores de section; la première, à savoir le coefficient alpha de Cronbach, renseigne sur la représentativité des items utilisés par rapport à l'univers possible des items mesurant le même trait, alors que la seconde que nous avons appelé corrélation inter-profil se base sur la définition de l'erreur de mesure que nous avons fournie précédemment.

Le tableau 3 présente ces deux séries de résultats pour deux scores d'insatisfaction, à savoir les MAL et MAB. Il est à noter que les scores de section ne sont pas calculés dans le cas des mesures de réalité et d'attentes.

L'examen de ce tableau fait ressortir l'étroit parallélisme entre la MAL et la MAB sur le plan de la fidélité. Au point de vue représentativité des items, on remarque que tous les coefficients alpha sont élevés; seul le coefficient de la section 6 est inférieur à 0.90. Considérant toutefois que ce score de section n'est basé que sur 5 items, le coefficient obtenu est très acceptable.

L'examen des corrélations inter-profils fait ressortir un phénomène très intéressant; en effet, on remarque qu'il existe un écart marqué entre la

Tableau 3

Fidélité des MAL et des MAB de section

Section	Nombre d'item	MAL alpha	MAL r^1	MAB alpha	MAB r^1
1	12	93	80	93	80
2	12	92	84	92	82
3	10	90	85	89	86
4	12	93	73	93	76
5	10	90	84	91	66
6	5	81	77	81	77
Score global	61	99	-	99	-

[1] Cette corrélation a été corrigée par la formule de Spearman-Brown.

fidélité de la MAL et celle de la MAB en ce qui a trait à la section 5. Cet écart est explicable par le fait que, cette section étant celle du test où l'insatisfaction est la moins élevée, il est plus probable qu'un léger écart A - B donne lieu à une variation des MAB qu'à une variation des MAL. Prenons à titre d'exemple une classe où deux étudiants sont insatisfaits, le premier indiquant une différence A - B = +1, et le second A - B = -1, et où le score vrai du professeur est 0.0, tant pour la MAB que pour la MAL. Supposons également que ces deux étudiants sont une erreur de mesure; leur effet sur la MAL est d'augmenter la variance intra-classe, la variance inter-classe demeurant inchangée. Leur effet sur la MAB est d'augmenter la variance intra-classe tout en diminuant la variance inter-classe, ce qui produit par définition une MAB moins fidèle que la MAL. Cette hypothèse devra cependant être vérifiée en approfondissant la logique statistique des scores de différence, point sur lequel personne ne semble encore avoir apporté d'éclaircissement majeur.

Ces corrélations inter-profils montrent donc une précision relativement élevée des scores de section.

Stabilité des coefficients de précision

Connaissant l'ordre de grandeur des coefficients de précision des items du test, il est d'égale importance de connaître la stabilité de ces coefficients. Quoique nous puissions théoriquement déduire cette stabilité du grand nombre d'observations utilisées pour estimer ces coefficients[1], il nous apparaît plus utile compte tenu de la fluctuation échantillonnale d'aller vérifier empiriquement la stabilité échantillonnale des coefficients de fidélité (précision).

Pour ce faire, une deuxième série de coefficients de fidélité a été calculée à partir d'un second échantillon de 728 classes, à savoir les professeurs ayant utilisé le test PEPT au printemps 1971; ces résultats font l'objet du tableau 4.

La comparaison de ces données à celles obtenues sur le premier échantillon se fera d'une part par l'examen des moyennes et des écarts-types des coefficients de chaque mesure entre les coefficients observés pour chaque échantillon; ces résultats font l'objet du tableau 5.

L'examen de ce tableau fait ressortir une identité quasi parfaite entre les deux séries de coefficients pour chaque mesure. A titre d'illustration, la fidélité moyenne des 61 items (mesure TM) est de 0.88 comparativement à 0.87 pour l'échantillon B. La dispersion de ces coefficients est de 0.05 dans chaque cas et la corrélation entre les deux échantillons (N = 61) est de 0.98.

Il semble donc évident que les coefficients obtenus précédemment fournissent une image très précise de la fidélité réelle des items du test.

---

[1] Les limites de confiance d'un coefficient de stabilité sont à la fois fonction de l'ordre de grandeur de ce coefficient et du nombre d'observations sur lequel il est basé (Lord et Novick, 1968, pp. 206-208).

Tableau 4

Coefficients de précision (échantillon B: 728 classes)

	M_0	M_1	M_2	M_3	M_4
1	.95	.80	.88	.78	.79
2	.80	.55	.67	.78	.72
3	.86	.66	.80	.85	.79
4	.86	.57	.81	.82	.76
5	.91	.82	.85	.85	.78
6	.87	.66	.83	.83	.76
7	.83	.66	.79	.80	.69
8	.85	.68	.79	.78	.69
9	.94	.90	.85	.75	.69
10	.91	.86	.68	.72	.67
11	.86	.78	.65	.64	.61
12	.90	.83	.80	.74	.69
13	.91	.84	.75	.74	.67
14	.90	.79	.80	.81	.71
15	.77	.55	.59	.76	.71
16	.8[illegible]	.69	.84	.84	.78
17	.89	.[illegible]	.79	.79	.73
18	.92	.81	.86	.85	.80
19	.95	.88	.87	.70	.65
20	.97	.91	.90	.83	.78
21	.98	.95	.94	.94	.91
22	.90	.78	.84	.80	.72
23	.88	.66	.80	.79	.71
24	.93	.87	.81	.86	.78
25	.90	.78	.85	.87	.81
26	.92	.79	.85	.81	.75
27	.89	.71	.81	.81	.75
28	.91	.87	.90	.86	.83
29	.98	.62	.93	.93	.89
30	.92	.77	.89	.89	.83
31	.96	.55	.92	.92	.87
32	.95	.78	.90	.89	.85
33	.95	.73	.86	.85	.79
34	.88	.62	.83	.83	.73
35	.80	.58	.86	.85	.76
36	.87	.58	.84	.85	.77
37	.82	.66	.71	.77	.67
38	.96	.70	.85	.88	.75
39	.87	.52	.86	.87	.76
40	.86	.57	.88	.88	.73
41	.84	.52	.86	.83	.71
42	.87	.59	.82	.83	.76
43	.[illegible]8	.55	.88	.85	.77
44	.1	.63	.68	.73	.65
45	.86	.[illegible]	.77	.75	.63
46	.81	.58	.76	.76	.66
47	.8[illegible]	.67	.85	.81	.75
48	.82	.55	.73	.77	.69
49	.86	.62	.81	.83	.77
50	.89	.89	.87	.87	.85
51	.85	.55	.82	.83	.75
52	.77	.83	.81	.86	.82
53	.92	.76	.87	.87	.81
54	.89	.75	.74	.78	.72
55	.87	.71	.86	.78	.70
56	.95	.83	.82	.63	.56
57	.89	.70	.83	.83	.77
58	.87	.71	.78	.76	.70
59	.87	.69	.78	.69	.64
60	.87	.62	.79	.77	.71
61	.85	.59	.75	.78	.68

Tableau 5

Moyennes et écarts-types des coefficients de précision et corrélation inter-échantillon

	ÉCHANTILLON A[1]		ÉCHANTILLON B[2]		
Mesure	Moy[3]	σ	Moy[3]	σ	$R_{A,B}$
MA	.88	.05	.90	.05	.98
MD	.68	.13	.74	.13	.96
MAL	.81	.07	.82	.08	.97
MAS	.80	.07	.81	.07	.97
%	.74	.07	.75	.07	.96

[1] 1196 classes

[2] 726 classes

[3] la transformation Z de Fisher a été utilisée.

Il est possible de tirer cette conclusion sans effectuer de test statistique puisque nous utilisons de fait des populations et non des échantillons.

Cette étude a été reprise en utilisant, cette fois 6 échantillons différents (Brunet, 1973), et les résultats que nous avons présentés ci-dessus se sont vérifiés dans chaque cas.

Nouveaux contrôles et implications
quant à la fidélité des mesures

Dans l'optique de la définition de l'erreur de mesure que nous avons donnée précédemment, il est évident que toute transformation des résultats bruts qui aurait pour effet de réduire la variance intra-classe sans modifier la variance inter-classes augmenterait la fidélité des mesures. Examinons brièvement quelques contrôles qui sont actuellement appliqués sur les données brutes, ainsi que quelques contrôles qu'il serait possible d'introduire.

A. Contrôles actuels

Les données brutes du questionnaire PERPE sont actuellement soumises à un contrôle qui consiste à identifier, dans une classe, un ou des étudiants marginalement insatisfaits, et à éliminer ces étudiants des calculs subséquents. Ce contrôle est décrit en détail dans Bernier et al. (1970).

L'augmentation de la fidélité des mesures suite à ce contrôle a été démontrée en partie (Gagné, 1973): en effet, l'élimination des cas marginaux a amené une hausse significative de la fidélité des MA (réalité) tandis que la hausse au niveau des mesures d'insatisfaction n'est pas significative (P<.50).

Le deuxième contrôle appliqué aux données brutes consiste à éliminer les étudiants qui n'ont pas répondu à un nombre suffisant d'items. Aucune donnée n'est disponible pour en évaluer l'effet sur les fidélités. Cependant, ce contrôle n'en est pas un qui touche le coeur du sujet, à savoir la réduction de la variance intra-classe.

B. Contrôles possibles

De nouveaux contrôles pourraient être introduits; ces contrôles permettraient d'expliquer et de contrôler certains facteurs étudiants, de sorte que la variance intra-classe soit réduite d'autant.

Deux types majeurs de contrôles ont jusqu'ici retenu notre attention, à savoir la sévérité différentielle des étudiants et le contrôle de certaines

caractéristiques "sociologiques" des étudiants.

1. La sévérité différentielle

Lorsqu'un étudiant essaie de situer un professeur par rapport à un item sur une échelle de réalité (échelle A), il le fait par rapport à un point de référence qui lui est propre. L'écart prononcé entre les points de référence des étudiants d'une classe est relié à la sévérité de chaque étudiant.

Advenant la mise au point d'une méthode qui permette d'obtenir pour chaque étudiant un indice de sa sévérité propre, il serait alors possible de corriger leurs réponses afin de ramener tous les étudiants de la classe à un même point de référence, réduisant la variance intra-classe.

2. Les caractéristiques "sociologiques"

La perception qu'a un étudiant de son professeur peut être explicable en partie par des caractéristiques "sociologiques" de l'étudiant, telles son sexe, son âge, son niveau académique, etc. Si tel était le cas, il y aurait lieu de contrôler ces variables en corrigeant les réponses des étudiants.

Les études menées par l'IREE-Education ont montré que l'influence de ce type de variable indépendante avaient une influence minime sur les réponses des étudiants (Beaudry, 1973) et que, par conséquent, toute correction des réponses étudiantes s'avérait inutile.

Dans une étape ultérieure, nous croyons qu'il serait possible d'étudier l'influence de caractéristiques étudiantes plus pertinentes au phénomène de la perception, à savoir les attitudes, etc.

C. Conclusion

En guise de conclusion, qu'il suffise de mentionner que l'identification de tout déterminant de la perception des étudiants ou de leur insatisfaction pourrait donner lieu à une correction des réponses de ces étudiants, correction qui permettrait de hausser la précision des mesures.

Résumé et conclusion

Après avoir défini d'une part la variance vraie comme étant la partie de la variance totale attribuable à la composante cours-professeur et, d'autre part la variance d'erreur comme étant la variance attribuable aux étudiants (variance intra-classe), le schème d'analyse de variance simple classes x étudiants a été appliqué sur 1196 classes dans lesquelles le questionnaire PERFE a été administré à l'automne 1970. La fidélité moyenne des items est de 0.90 pour les MA, 0.68 pour les MB, 0.83 pour les MAB, 0.82 pour les MAL et 0.76 pour le pourcentage des étudiants insatisfaits, une fois éliminés 12 items dont la fidélité est jugée trop faible. La fidélité des scores composés est acceptable, telle que mesurée par le coefficient alpha de Cronbach et par la technique du "split-half" sur les classes.

Les coefficients de fidélité obtenus sont stables, i.e. identiques sur deux échantillons différents de classes. Enfin, certains contrôles sur les données pourraient permettre d'augmenter la fidélité des mesures.

Appendice A

Fidélité intra-profil

Chronologiquement, les premiers résultats disponibles quant à la fidélité du questionnaire ont été ceux que nous avons appelé "fidélité intra-profil".

[illegible]

Etant donné Z, l'ensemble des étudiants d'une classe, deux sous-groupes sont formés au hasard (X et Y). Deux profils sont calculés, le premier à partir des réponses du sous-groupe X et le second à partir de celles du sous-groupe Y. Cette procédure est appliquée à chaque classe de l'échantillon, de sorte que, pour chaque classe, deux vecteurs de MA, de MB, de MAL et de MAB calculés sur des demi-classes sont disponibles.

Puis, pour chaque classe, la corrélation entre les deux vecteurs, pour chaque mesure, est calculée ($R_{MA(X)}$, MA(Y), $R_{MB(X)}$, MB(Y), $R_{MAB(X)}$, MAB(Y) et $R_{MAL(X)}$, MAL(Y)). Pour chaque mesure, la distribution des corrélations, ainsi que leurs paramètres, serviront d'indices de fidélité intra-profil et de cohésion des juges.

Puisque chaque sous-groupe X et Y a été formé au hasard, il nous semble raisonnable de postuler que le sous-groupe Y est équivalent au sous-groupe X. Par conséquent, les différences entre les profils tracés par chaque sous-groupe sont causés à la fois par l'inconsistance des juges à l'intérieur de chaque sous-groupe et par l'instabilité de la situation mesurée.

L'inférence quant à la stabilité temporelle intra-profil peut être faite à la condition de poser un second postulat, à savoir que, étant donné l'équivalence des deux sous-groupes X et Y à l'intérieur de chaque classe, le profil tracé par le sous-groupe X constitue le test alors que celui tracé par le sous-groupe Y constitue le retest, les conditions expérimentales, la situation mesurée et l'état des répondants étant maintenus constants.

En d'autres mots, ce schème permettra de dire jusqu'à quel point, lors de deux passations différentes du test par un même groupe, un professeur

recevrait un portrait pédagogique à l'intérieur duquel les items les plus gratifiants demeureraient toujours les plus gratifiants, alors que ceux qui le sont le moins le seraient toujours, moyennant qu'il n'y ait pas eu évolution de la situation mesurée ou des répondants.

Critique méthodologique[1]

Cette méthodologie a le defaut d'être sensible au fait que les moyennes des items soient différentes; en effet, dans le cas extrêmes de 5 items dont les moyennes sont différentes et dont le recouvrement des distributions est nul, la corrélation sera toujours egale à 1.0. Cependant, les items du questionnaire PERPE sont loin de ressembler à cet exemple extrême; en effet, pour la plupart des items, les moyennes sont sensiblement identiques, de sorte que l'information fournie par cette méthodologie, quoique contaminée en partie par les écarts entre les moyennes, peut à notre avis éclairer et compléter les résultats obtenus par l'analyse de variance.

Résultats

Les distributions cumulées des corrélations intra-profils des MA et MB font l'objet du tableau 6, tandis que celles des MAL et MAB sont présentées au tableau 7. Ces résultats sont représentés graphiquement par les figures. Si on ne conserve que les classes de 12 étudiants et plus, les moyennes de ces corrélations sont de .92, .90, .74 et .81 pour les MA, MB, MAB et MAL.

Les distributions présentées aux tableaux 6 et 7 ne tiennent pas compte d'une variable importante, à savoir le nombre d'étudiants par classe; en effet, le nombre d'étudiants dans la classe influence directement la fidélité des profils, de sorte qu'un profil basé sur 30 étudiants est probablement plus fidèle qu'un profil basé sur 10 étudiants. En conséquence, la corrélation intra-profil risque d'être plus élevée pour les classes où le

---

[1] La critique a été formulée par Denis Allaire.

Tableau 6

Distribution des corrélations des MA et MB

MA			MB		
R	FC[1]	FCP[2]	R	FC[1]	FCP[2]
0,99	1906	1,00	0,98	1906	1,00
0,97	1905	0,99	0,96	1904	0,99
0,95	1891	0,99	0,94	1867	0,98
0,93	1730	0,91	0,92	1787	0,93
0,91	1473	0,77	0,90	1581	0,82
0,89	1153	0,60	0,88	1281	0,67
0,87	820	0,43	0,86	964	0,50
0,85	587	0,30	0,84	696	0,36
0,83	394	0,20	0,82	470	0,24
0,81	281	0,14	0,80	303	0,15
0,79	193	0,10	0,78	186	0,09
0,77	122	0,06	0,76	120	0,06
0,75	72	0,03	0,74	74	0,03
0,73	57	0,02	0,72	53	0,02
0,71	36	0,01	0,70	33	0,01
0,69	25	0,01	0,68	29	0,01
0,67	20	0,01	0,66	21	0,01
0,65	13	0,00	0,64	17	0,00
0,63	8	0,00	0,62	10	0,00
0,61	6	0,00	0,60	9	0,00
0,59	2	0,00	0,58	6	0,00
0,57	2	0,00	0,56	2	0,00
0,55	2	0,00	0,54	2	0,00
0,53	1	0,00	0,52	2	0,00
0,51	1	0,00	0,50	1	0,00
			0,48	1	0,00
			0,46	1	0,00
			0,44	1	0,00
			0,42	1	0,00

[1] Fréquences cumulées

[2] Fréquences cumulées en proportion

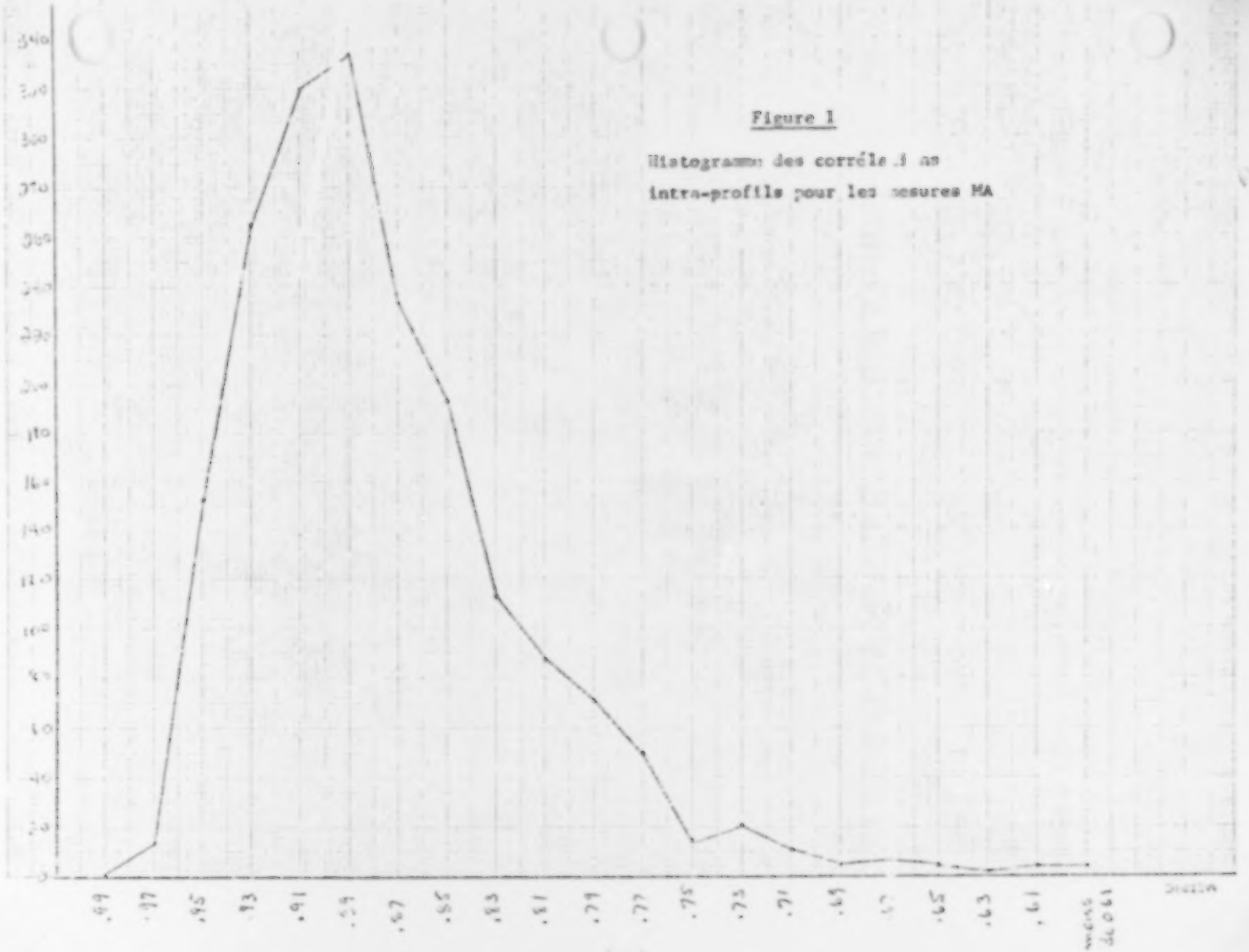

Figure 1

Histogramme des corrélations intra-profils pour les mesures MA

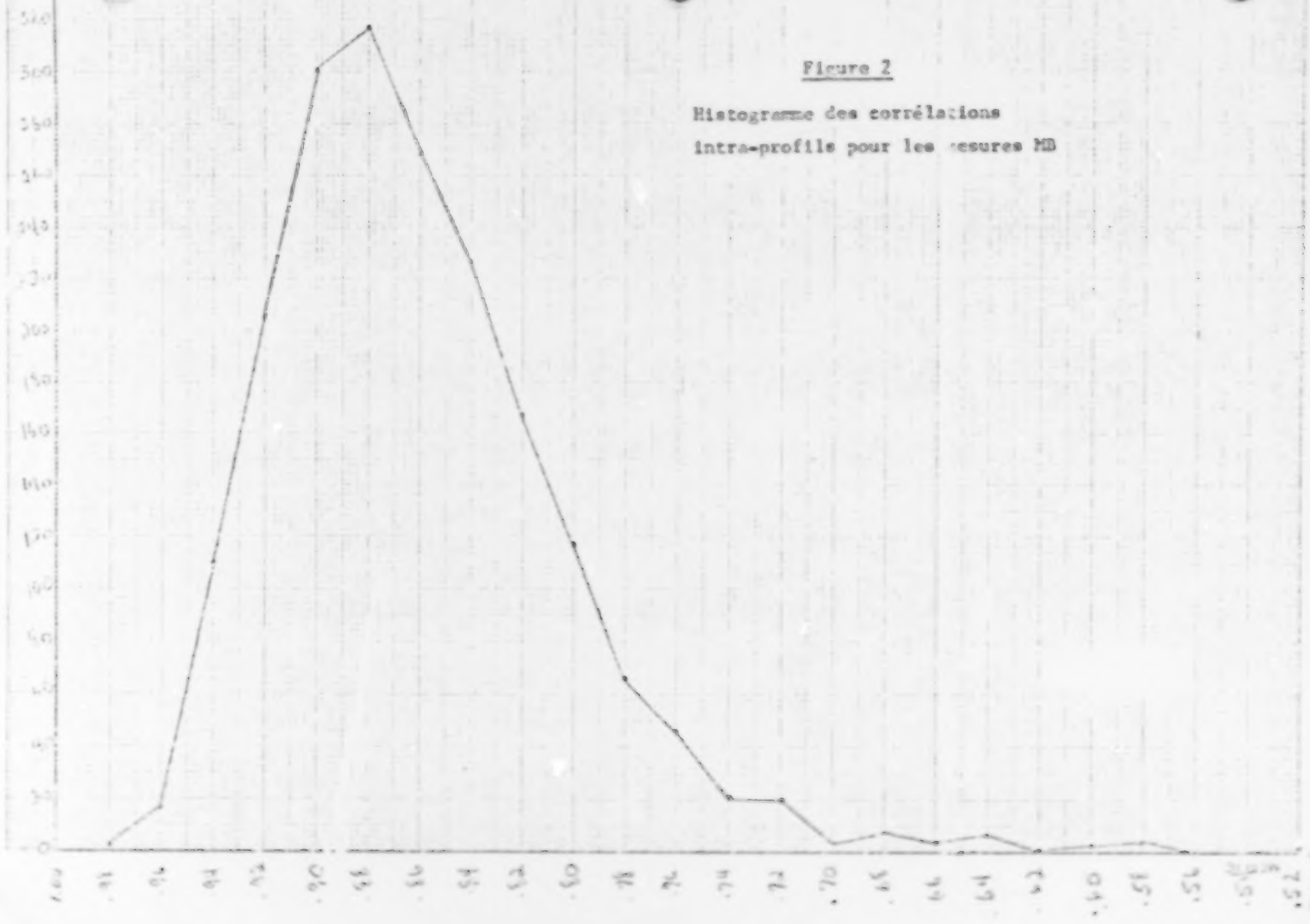

Figure 2

Histogramme des corrélations
intra-profile pour les mesures MB

Tableau 7

Distribution des corrélations des MAB et MAL

MAB			MAL		
R	FC[1]	FCP[2]	R	FC[1]	FCP[2]
0,92	1906	1,00	0,96	1906	1,00
0,89	1900	0,99	0,93	1905	0,99
0,86	1864	0,97	0,90	1864	0,97
0,83	1790	0,93	0,87	1740	0,91
0,80	1673	0,87	0,84	1566	0,82
0,77	1512	0,79	0,81	1355	0,71
0,74	1330	0,69	0,78	1114	0,58
0,71	1138	0,59	0,75	917	0,48
0,68	946	0,49	0,72	719	0,37
0,65	792	0,41	0,69	544	0,28
0,62	646	0,33	0,66	411	0,21
0,59	520	0,27	0,63	302	0,15
0,56	414	0,21	0,60	227	0,11
0,53	322	0,16	0,57	173	0,09
0,50	257	0,13	0,54	132	0,06
0,47	187	0,09	0,51	96	0,05
0,44	152	0,07	0,48	70	0,03
0,41	115	0,06	0,45	52	0,02
0,38	86	0,04	0,42	39	0,02
0,35	65	0,03	0,39	32	0,01
0,32	51	0,02	0,36	21	0,01
0,29	38	0,01	0,33	15	0,00
0,26	28	0,01	0,30	12	0,00
0,23	24	0,01	0,27	11	0,00
0,20	17	0,00	0,24	10	0,00
0,17	11	0,00	0,21	8	0,00
0,16 et moins	8	0,00	0,20 et moins	6	0,00

[1] Fréquences cumulées

[2] Fréquences cumulées en proportion

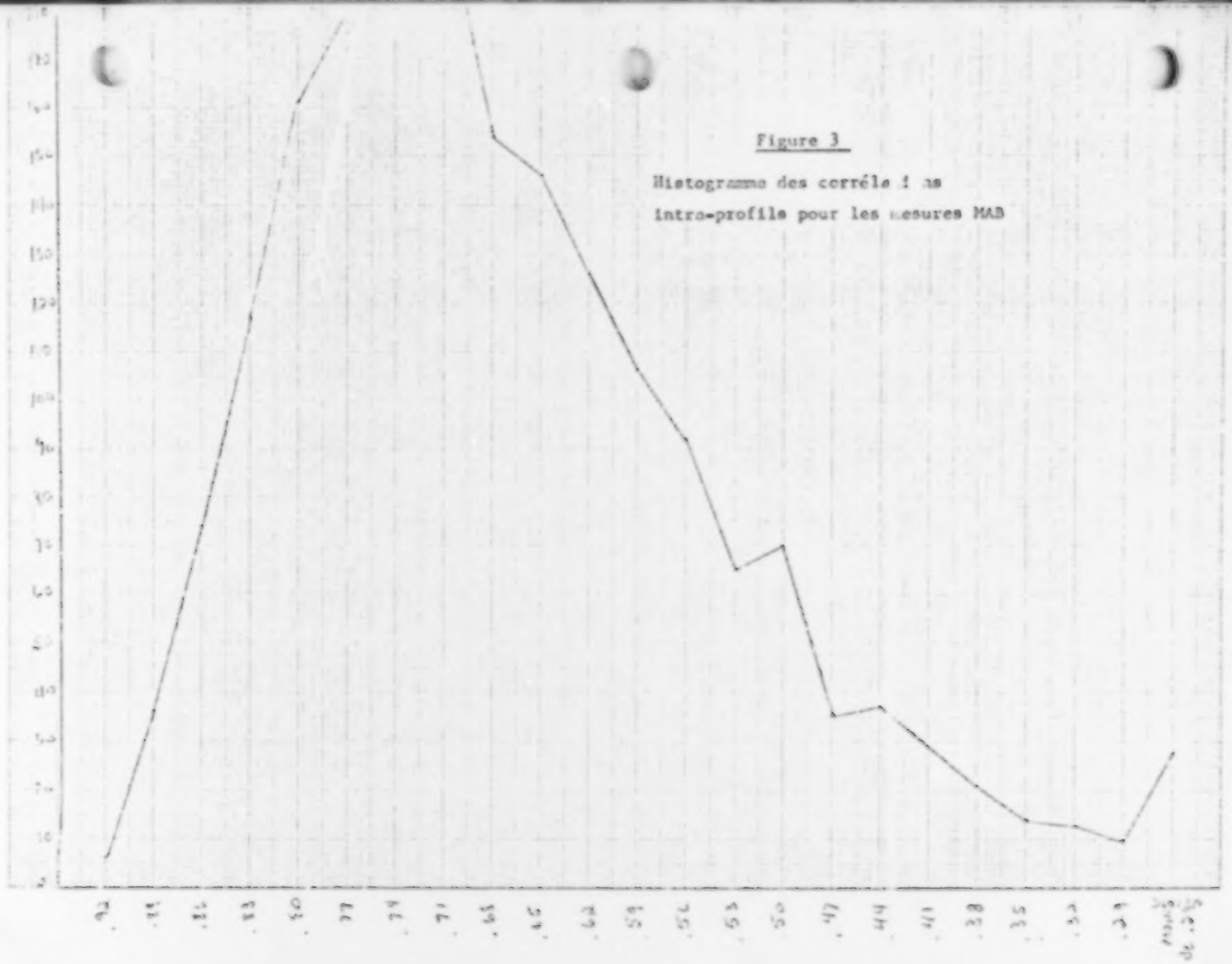

Figure 3

Histogramme des corrélations intra-profile pour les mesures MAB

Figure 4

Histogramme des corrélations
intra-profils pour les mesures MAL

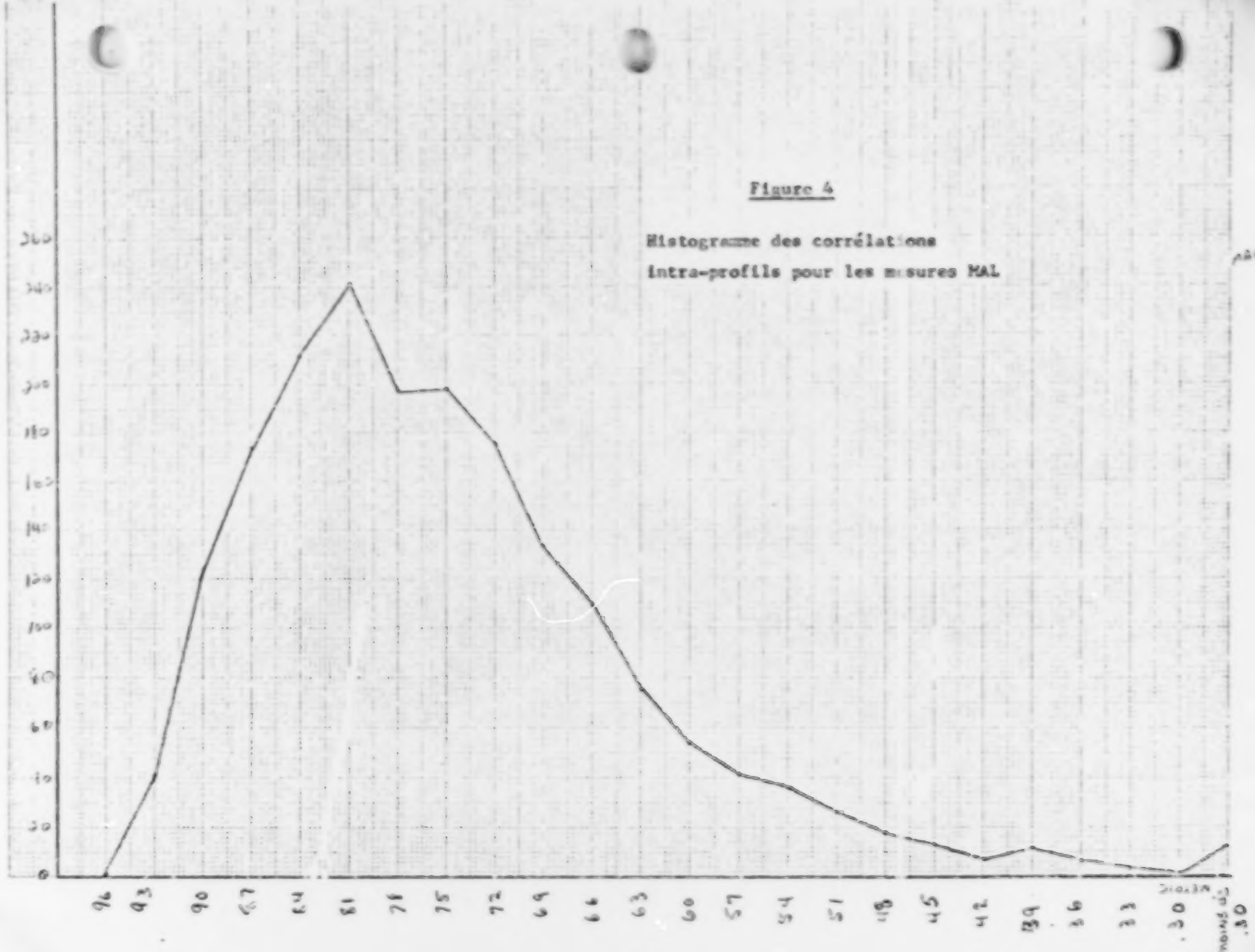

Tableau 8

Corrélations moyennes observées
en fonction du nombre de répondants par classe

NR	$\bar{R}_{MA}$	N_{MA}	$\bar{R}_{MB}$	N_{MB}	$\bar{R}_{MAB}$	N_{MAB}	$\bar{R}_{MAL}$	N_{MAL}
7	.85	96	.83	9[illegible]	.60	96	.66	96
8	.87	155	.85	157	.63	157	.71	157
9	[illegible]	173	.86	169	.64	171	.72	171
10	.90	187	.88	190	.68	190	.76	190
11	.90	249	.89	246	.69	246	.77	246
12	.91	226	.90	225	.72	226	.80	226
13	.92	201	.91	204	.73	203	.81	203
14	.93	172	.91	172	.75	171	.83	171
15	.94	124	.92	122	.75	125	.82	125
16	.94	93	.93	97	.77	93	.84	93
17	.95	59	.92	58	.79	59	.85	59
18	.95	48	.93	47	.80	47	.87	47
19	.95	36	.94	36	.81	35	.87	35
20	.97	18	.95	19	.83	18	.89	18
21	.96	11	.93	11	.82	11	.87	11
22	.96	9	.95	9	.84	10	.87	10
23	.96	8	.95	8	.83	7	.89	7
24	.95	9	.94	9	.86	9	.90	9
25	.97	3	.98	3	.80	3	.94	3
26	-	0	-	0	-	0	-	0
27	.97	29	.96	29	.86	29	.92	29

nombre d'étudiants est élevé. Le tableau 8 montre l'effet du nombre d'étudiants sur les corrélations intra-profils. On remarque une croissance continue des moyennes en fonction du nombre d'étudiants dans chaque demi-classe.

Le tableau 9 montre, pour des demi-classes de grandeur différente,

Tableau 9

Limites inférieures des corrélations observées
en fonction du nombre de répondants par classe

N. Rép.	MA	MB	MAB	MAL
7	.55	.55	•	•
8	.65	.60	•	•
9	.70	.65	•	•
10	.70	.65	•	•
11	.70	.65	•	•
12	.75	.75	•	•
13	.70	.75	•	•
14	.80	.75	•	•
15	.80	.80	•	•
16	.80	.80	•	•
17	.80	.85	•	.65
18	.85	.85	.65	.75
19	.85	.85	.55	.65
20	.90	.90	.70	.80
21	.90	.90	.70	.75
22	.90	.85	.75	.80
23	.90	.90	.70	.85
24	.90	.85	.70	.80
25	.95	.95	.75	.[illegible]
26	-	-	-	-
27	.95	.90	.70	.80

• Inférieur à 0.50

Tableau 10

Corrélation au dessus de laquelle se situe un % donné de sous-groupes

Mesure	90%	80%	70%	60%	50%
MA	.80	.83	.85	.86	.88
MB	.80	.82	.83	.85	.86
[illegible]	.50	.56	.61	.64	.69
MAL	.60	.66	.70	.74	.77

la limite inférieure observée de la corrélation intra-profil. On remarque encore la croissance de cette limite inférieure en fonction du nombre d'étudiants dans la demi-classe.

Tous ces résultats sont synthétisés au tableau 10 qui présente pour chaque mesure et pour divers centiles la corrélation intra-profil observée. Ainsi, on voit que, au niveau de la MA, 90% des corrélations observées sont égales ou supérieures à .80.

Appendice B

Fidélité inter-profil

Méthodologie

Ayant pour chaque classe les profils tracés par les sous-groupes X et Y, il est possible de calculer pour chaque item sur l'ensemble des classes la corrélation entre la mesure calculée à partir de X et de Y. On obtient ainsi une corrélation par item et cette corrélation, corrigée par la formule de Spearman Brown, constitue un indice de la fidélité de cet item. En effet, ce coefficient reflète jusqu'à quel point l'ordre des professeurs à un item donné est conservé d'une administration à l'autre, moyennant évidemment qu'on accepte les postulats posés à l'appendice A.

Commentaire méthodologique

Il peut être démontré que la technique d'analyse de variance utilisée dans ce rapport a un rationel tout à fait identique à la technique décrite plus haut. Algébriquement, le coefficient obtenu par ANOVA est égal à la valeur attendue des coefficients obtenus par le "split-half" des classes, ce "split-half" au hasard étant repris un grand nombre de fois.

Echantillon

Les coefficients ont été calculés à partir de trois échantillons de données:

échantillon A_1 : 575 classes d'au moins 20 étudiants (administration à l'automne 1970)

échantillon A_2 : 572 classes d'au moins 24 étudiants (administration à l'automne 1970)

échantillon B : 125 classes d'au moins 24 étudiants (administration au printemps 1971).

Résultats

Le tableau 11 contient, pour chaque item, chaque mesure et chaque

Tabl[illegible] 1

Corrélations inter-profils des dimensions pour trois échantillons[1]

Dim.	Echantillon A_1				Echantillon A_2				Echantillon B			
	MA	MB	MAL	MAB	MA	MB	MAL	MAB	MA	MB	MAL	MAB
01	0.91	0.77	0.82	0.72	0.93	0.83	0.85	0.80	0.90	0.75	0.33	0.61
02	0.74	0.57	0.59	0.67	0.77	0.65	0.60	0.71	0.76	0.50	0.54	0.77
03	0.79	0.50	0.69	0.74	0.83	0.60	0.72	0.76	0.80	0.59	0.76	0.85
04	0.79	0.58	0.69	0.71	0.84	0.70	0.75	0.75	0.85	0.54	0.82	0.83
05	0.84	0.72	0.76	0.76	0.87	0.81	0.77	0.77	0.83	0.74	0.70	0.71
06	0.79	0.58	0.72	0.72	0.84	0.73	0.76	0.76	0.81	0.53	0.77	0.77
07	0.72	0.56	0.63	0.67	0.76	0.67	0.66	0.72	0.76	0.40	0.61	0.65
08	0.74	0.60	0.60	0.63	0.77	0.73	0.67	0.66	0.7[illegible]	0.37	0.64	0.67
09	0.89	0.79	0.78	0.66	0.90	0.82	0.78	0.68	0.[illegible]	0.85	0.70	0.70
10	0.85	0.75	0.58	0.54	0.87	0.76	0.61	0.58	0.8[illegible]	0.82	0.63	0.63
11	0.76	0.65	0.5[illegible]	0.48	0.80	0.70	0.59	0.51	0.[illegible]3	0.77	0.55	0.55
12	0.85	0.75	0.6[illegible]	0.65	0.84	0.76	0.66	0.66	0.[illegible]4	0.68	0.67	0.65
13	0.85	0.72	0.69	0.64	0.85	0.75	0.72	0.64	0.83	0.75	0.66	0.55
14	0.80	0.62	0.69	0.70	0.82	0.68	0.70	0.69	[illegible].83	0.67	0.62	0.65
15	0.59	0.44	0.[illegible]4	0.59	0.66	0.55	0.55	0.67	[illegible].65	0.47	0.54	0.65
16	0.85	0.60	0.[illegible]3	0.74	0.85	0.67	0.73	0.76	[illegible].88	0.64	0.78	0.77
17	0.83	0.69	0.69	0.69	0.84	0.71	0.70	0.70	0.[illegible]5	0.72	0.66	0.64
18	0.85	0.69	0.77	0.72	0.86	0.72	0.76	0.74	0.86	0.77	0.79	0.79
19	0.90	0.79	0.75	0.49	0.91	0.80	0.75	0.44	0.91	0.82	0.74	0.54
20	0.96	0.85	0.86	0.74	0.95	0.86	0.84	0.75	0.95	0.86	0.86	0.80
21	0.97	0.88	0.88	0.88	0.97	0.88	0.89	0.89	0.97	0.93	0.92	0.91
22	0.82	0.59	0.71	0.69	0.83	0.67	0.72	0.70	0.88	0.72	0.72	0.66
23	0.83	0.59	0.70	0.65	0.85	0.69	0.73	0.67	0.89	0.63	0.74	0.76
24	0.87	0.75	0.71	0.69	0.88	0.76	0.73	0.70	0.90	0.80	0.67	0.64
25	0.82	0.66	0.70	0.75	0.81	0.70	0.67	0.77	0.87	0.73	0.75	0.74

Tableau (suite)

Dim.	Echantillon A_1				Echantillon A_2				Echantillon B			
	MA	MB	MAL	MAB	MA	MB	MAL	MAB	MA	MB	MAL	MAB
26	0.88	0.72	0.72	0.64	0.89	0.79	0.74	0.68	0.67	0.63	0.83	0.78
27	0.85	0.67	0.72	0.72	0.88	0.78	0.75	0.75	0.66	0.60	0.68	0.69
28	0.89	0.59	0.87	0.81	0.91	0.72	0.89	0.84	0.00	0.32	0.86	0.77
29	0.92	0.68	0.89	0.89	0.93	0.79	0.89	0.90	0.03	0.53	0.89	0.90
30	0.88	0.71	0.83	0.84	0.90	0.80	0.84	0.86	0.06	0.51	0.79	0.81
31	0.91	0.65	0.87	0.87	0.92	0.77	0.89	0.89	0.92	0.36	0.88	0.88
32	0.91	0.65	0.83	0.81	0.92	0.75	0.85	0.83	0.93	0.73	0.86	0.85
33	0.88	0.62	0.80	0.78	0.89	0.66	0.78	0.79	0.93	0.72	0.83	0.82
34	0.83	0.57	0.72	0.75	0.85	0.66	0.74	0.76	0.85	0.59	0.75	0.77
35	0.80	0.63	0.70	0.71	0.85	0.75	0.69	0.73	0.7?	0.30	0.76	0.78
36	0.80	0.64	0.74	0.74	0.83	0.76	0.72	0.74	0.81	0.39	0.75	0.75
37	0.70	0.51	0.5?	0.62	0.72	0.59	0.56	0.65	0.?7	0.49	0.60	0.59
38	0.85	0.61	0.??	0.76	0.86	0.68	0.76	0.78	0.87	0.57	0.78	0.78
39	0.81	0.58	0.76	0.76	0.82	0.72	0.77	0.76	0.78	0.34	0.77	0.76
40	0.77	0.62	0.68	0.67	0.79	0.71	0.71	0.69	0.80	0.46	0.74	0.75
41	0.75	0.57	0.67	0.68	0.80	0.71	0.72	0.73	0.76	0.46	0.70	0.70
42	0.78	0.48	0.65	0.65	0.81	0.57	0.74	0.73	0.81	0.38	0.78	0.77
43	0.81	0.59	0.76	0.76	0.81	0.69	0.77	0.78	0.86	0.36	0.80	0.80
44	0.71	0.59	0.56	0.63	0.76	0.70	0.59	0.66	0.77	0.60	0.56	0.64
45	0.80	0.64	0.62	0.60	0.82	0.68	0.66	0.65	0.85	0.63	0.72	0.57
46	0.69	0.49	0.60	0.61	0.72	0.59	0.58	0.61	0.72	0.45	0.59	0.59
47	0.82	0.61	0.75	0.72	0.84	0.69	0.73	0.77	0.79	0.53	0.70	0.60
48	0.66	0.43	0.59	0.64	0.68	0.53	0.61	0.67	0.79	0.53	0.66	0.67
49	0.78	0.66	0.68	0.70	0.85	0.76	0.75	0.76	0.82	0.45	0.68	0.76
50	0.82	0.61	0.75	0.76	0.83	0.70	0.77	0.78	0.79	0.38	0.78	0.78

Tableau (suite)

Dim.	Echantillon A_1				Echantillon A_2				Echantillon B			
	MA	MB	MAL	MAB	MA	MB	MAL	MAB	MA	MB	MAL	MAB
51	0.73	0.59	0.61	0.63	0.79	0.70	0.67	0.70	0.77	0.38	0.70	0.70
52	0.67	0.38	0.71	0.79	0.73	0.51	0.74	0.82	0.70	0.39	0.73	0.83
53	0.87	0.68	0.80	0.78	0.89	0.77	0.83	0.81	0.90	0.69	0.85	0.85
54	0.81	0.61	0.59	0.62	0.85	0.71	0.61	0.65	0.86	0.70	0.72	0.72
55	0.80	0.58	0.67	0.65	0.84	0.67	0.71	0.66	0.74	0.54	0.65	0.76
56	0.87	0.73	0.68	0.44	0.90	0.74	0.72	0.49	0.93	0.74	0.77	0.55
57	0.83	0.63	0.71	0.71	0.84	0.74	0.73	0.72	0.88	0.64	0.78	0.75
58	0.79	0.61	0.62	0.63	0.81	0.68	0.66	0.66	0.85	0.70	0.69	0.67
59	0.80	0.55	0.53	0.53	0.81	0.64	0.57	0.57	0.81	0.66	0.56	0.54
60	0.78	0.52	0.63	0.60	0.79	0.61	0.62	0.60	0.81	0.51	0.66	0.63
61	0.74	0.45	0.58	0.58	0.74	0.57	0.62	0.63	0.77	0.55	0.69	0.64

---

[1] Echantillon A_1 : 575 classes d'au moins 20 étudiants, production A (automne 1970)

Echantillon A_2 : 312 classes d'au moins 24 étudiants, production A (automne 1970)

Echantillon B : 125 classes d'au moins 24 étudiants, production B (printemps 1971)

N.B. chaque classe a été divisée en 2 sous-groupes égaux.

échantillon, la corrélation corrigée par la formule de Spearman Brown[1].

Comme l'interprétation de ces résultats risque fort d'être redondante par rapport à la troisième partie de ce document, nous suggérons au lecteur de lire cette partie.

---

1 $r_{tt} = \frac{2\, r_o}{1 + r_o}$

RÉFÉRENCES

BEAUDRY, M. (1973) L'influence des caractéristiques "étudiant" et "professeur" sur les mesures d'attente, de réalité et d'insatisfaction du test PERPE. Communication présentée au 41ème congrès de l'ACFAS à Montréal. INRS-Education, document 0573-08.

BERNIER, L. et al. (1970) Test PERPE: Guide d'analyse des résultats. INRS-Education.

BRUNET, R. (1973) Données sur la fidélité du questionnaire PERPE. Petit Rapport Psychométrique, no. 21.

GAGNE, F. (1970) Les six nouvelles sections du test PERPE. INRS-Education, document 0670-01.

GAGNE, F. (1973) Gains de fidélité grâce aux cas marginaux. Petit Rapport Psychométrique, no. 20.

HAMEL, C. (1972) Les sections du test PERPE. INRS-Education, Document 0572-04.

LORD, F.M. & NOVICK, M.R. (1968) Statistical theories of mental test scores. Reading: Addison-Wesley.

NUNNALLY, J.C. (1967) Psychometric Theory. New York: McGraw-Hill.

RENAUD, Micheline (1973) Rapport des statistiques d'utilisation: 1970-1972. (sous presse) INRS-Education.

STANLEY, J.C. (1971) Reliability. In Thorndike, R.L. (ed.) (1971) Educational measurement. Washington: ACE

WINER, B.J. (1971) Statistical principles in experimental design (2e ed.) New York: McGraw-Hill.